Concept Dev for COMP study

Sitaram Devarakonda & George Chacko

August 2, 2019

1 Motivation

Making discovery in a manner that would leverage existing resources and skills with focus on the computer science field, which has experienced vigorous growth in the last twenty years. Citation analysis is a particular focus.

2 Data and Methods

The following sections describe the datasets and initial data exploration

2.1 Year slices

Year slices were developed for all publications from 1996-2018. All publications of language 'English' and with at-least 2 references and complete publication data in Scopus are selected. Publications type include (article, conference paper, dissertation etc) and belong to subject area 'COMP'.

2.2 Disciplinary Composition

One way to identify the growth and diversity of a field might be the subject areas of publications and how it changes over time .So We calculated the disciplinary composition for all the publications and unique references from 1996-2018. These are calculated based on ASJC codes which are assigned by Scopus at journal level, i.e every publication in a journal or conference will have the same subject area. Journals or conferences can have multiple subject areas. Multidisciplinary journals such as Nature, Science, Proceedings of the National Academy of Sciences are classified as General which result in the loss of computer science articles published in those journals. Also there is a lot of ambiguity in journals/conferences with multiple ASJC codes. Following possibilities can be explored to overcome ambiguity:

Table 1: Summary of base Scopus Analytical data set. The number of unique publications, unique references, total references .

Year	Unique Publications	Unique References	Total References
1996	30,740	184,617	318,700
1997	37,166	228,464	386,627
1998	44,764	250,491	438,908
1999	40,434	237,175	413,681
2000	44,636	261,393	460,310
2001	62,000	343,937	619,330
2002	71,645	394,882	722,278
2003	83,676	438,498	849,950
2004	82,914	482,475	915,997
2005	107,905	591,784	1,193,198
2006	139,379	713,924	1,531,101
2007	170,396	829,444	1,830,944
2008	204,084	1,000,790	2,226,980
2009	239,113	1,222,312	2,790,194
2010	254,763	1,374,471	3,164,912
2011	255,144	1,519,064	3,419,719
2012	170,364	1,374,757	2,785,900
2013	195,008	1,606,955	3,353,390
2014	264,968	1,863,488	4,260,613
2015	297,534	2,037,838	4,762,036
2016	305,372	2,217,138	5,140,149
2017	291,481	2,349,259	5,262,725
2018	316,495	2,623,449	6,009,386

- NSF follows a single subject area for journals which can be used for publication level classification.
- Ludo Waltman methodology for publication-level classification (uses direct citation)
- Kevin Boyack methodology
- Using Scopus FPE to generate publication level classification
- Obtain top conferences/journals in COMP field and analyze their citation patterns. (Assuming they contain articles related to single discipline)

2.3 Possible next steps(based on discussion)

Analyze the citation patterns of top authors in the field and how it evolved over time. Selecting authors with at least 30 papers published might be a good place to start. Another experiment might be to approach the diversity of computer science from other disciplines. i.e studying the citation patterns of other disciplines and how computer science has been more inclusive over time.